局所特徴量を用いた構音障害者の音声認識の検討*

◎宮本千琴, 滝口哲也, 有木康雄 (神戸大・工), 李義昭 (追手門大), 中林稔堯 (神戸大・発達)

1 はじめに

近年，音声認識技術の発展に伴い，様々な環境下や場面での利用が期待されている．例えばカーナビゲーションの操作や会議音声の議事録化など様々な分野に応用されている．これらの多くは健常者を対象としており，文献 [1] [2] では，構音障害者音声を対象とした特徴量抽出や音響モデル適応，構築を行っているが，言語障害者などの障害者を対象としたものは非常に少ない．

言語障害の原因の一つとして，脳性麻痺が考えられる．意図的な動作を行う場合や緊張状態にある時に筋肉の制御が難しくなり，不随意運動を伴う．この運動障害の一つとして，正しく構音できない場合がある [3]．本稿では，脳性麻痺により構音機能に障害を持つ被験者を対象に音声認識実験を行う．

我々はこれまで，時間-周波数平面上の局所的な幾何学構造を抽出する方法を提案してきた [4] [5]．これらの手法は画像認識の分野においても有効性が示されている [6]．本稿では，音声の時間-デルタケプストラム平面上の局所的な幾何学構造に基づき特徴を抽出する手法を提案し，単語認識実験を行い，その有効性を示す．

2 局所特徴量

2.1 局所パターンと局所特徴量行列

局所特徴量とは，ある点周辺でのあるパターンの値の強さを表した特徴量のことである．Fig.1 の局所パターンを，時間-デルタケプストラム平面に適用したものが局所特徴量の一例である．

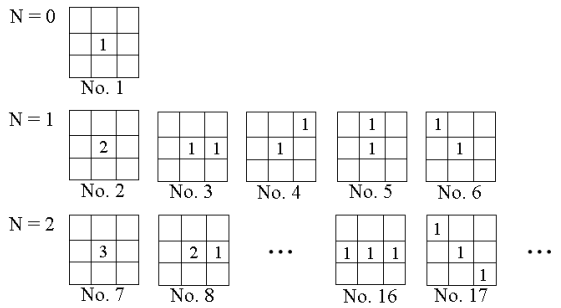

Fig. 1 局所パターンの例

時間-デルタケプストラム平面において点 $r(t,d)$ (時刻 $t = 1,\cdots,T$，デルタケプストラムの次数 $d = 1,\cdots,D$) でのデルタケプストラムの値を $I(r)$ とすると，点 r での局所パターン k の局所特徴量 $h_k(r)$ は次式で表される．

$$h_k(r) = I(r) + I(r + a_1^{(k)}) \cdots + I(r + a_N^{(k)}) \quad (1)$$

本稿では局所パターン k の変位 $(a_1^{(k)},\cdots,a_N^{(k)})$ を参照点 r の局所領域 3×3 に限定し，次数 N を高々2 までに制限すると，局所パターン K の数は平行移動により等価なものを除くと全部で 35 種類になる（Fig.1）．局所パターンの 1 に対応するデルタケプストラムの値を加算することにより，各々の局所パターンに対応する局所特徴量が得られる．ただし，2 は 2 倍，3 は 3 倍を意味する．

2.2 音声特徴量ベクトル

時間-デルタケプストラム平面上で得られた局所特徴量をフレーム内で縦につなげたベクトル $\mathbf{x}$ を音声特徴量ベクトルとする．時間 t におけるベクトルは以下のように M 次元ベクトル（$M = K \times (D-2)$）で表記する．

$$\mathbf{h^{(t)}} = [h_{1,2}^{(t)} \cdots h_{K,D-1}^{(t)}]^t \quad (2)$$

3 認識実験

3.1 実験条件

実験用データとして構音障害者，健常者それぞれ 1 名のデータを収録した．発話内容として ATR 音素バランス単語（216 単語）から 210 単語を無作為に選択した．収録は各単語を 5 回連続発声し，その後，各発話を手動で切り出した．収録データのサンプリング周波数は 16 kHz，フレーム窓長は 25 msec，フレーム周期は 10 msec であり，時間-デルタケプストラム平面として ΔMFCC（12 次元）を用いた．音響モデルは monophone（54 音素，8 混合）を用い，1 回目の発話の認識を行う場合は 2〜5 回目の発話を用いて作成した．これを各発話に対して行う．

3.2 構音障害者モデルでの認識実験

初めに，音声特徴量として，MFCC，ΔMFCC，MFCC+ΔMFCC を用いた認識実験を行った．5 回発話全体の平均の認識結果を表 1 に示す．

*A Study on Dysarthric Speech Recognition using Local Features, by Chikoto Miyamoto, Tetsuya Takiguchi, Yasuo Ariki (Kobe Univ.), Ichao Li (Otemon Univ.) and Toshitaka Nakabayashi (Kobe Univ.)

Table 1　特定話者モデルでの認識実験結果 [%]

特徴量	健常者	構音障害者
MFCC	99.2	78.5
ΔMFCC	98.4	49.2
MFCC+ΔMFCC	99.6	89.5

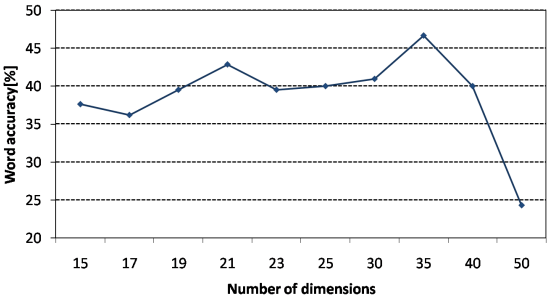

Fig. 2　次元数による認識率の推移（1 回目発話）

構音障害者の ΔMFCC が 49.2 %と健常者に比べると，著しく認識精度が低いことがわかる．これは，構音障害者の発話が健常者に比べて不安定になり，時間変化がうまく表現できていないと考えられる．そこで ΔMFCC の認識精度を向上させることで，時間特徴をよりよく表すことができる．

3.3　音声特徴ベクトルの次元圧縮

音声特徴量 **H** の次元数は，35（局所パターンの数）× 10（デルタケプストラムの次数 - 2）= 350 次元であり，高次元であることから，HMM の確率推定に問題が生じる可能性がある．そこで，音声特徴量 **H** を主成分分析（PCA）により次元圧縮を行う．1 回目の発話をテストデータとして用いた場合の次元数による認識率の変化を図 2 に示す．

結果より，35 次元以降は次元数の増加による認識率の改善が見られず，35 次元のときに 46.7 %と最も高い認識率が得られた．これより，以後の実験では 35 次元に圧縮したものを用いることとする．

3.4　提案手法での認識実験結果

時間-デルタケプストラム平面上の局所特徴量を PCA により 35 次元に圧縮した特徴量を用いて実験を行った．また、時間-ケプストラム平面として MFCC（12 次元），時間-周波数平面として対数メルフィルタバンク出力（24 次元）を用い，それぞれの平面上の局所特徴量を抽出し実験を行った．5 回発話全体の平均の結果を表 2 に示す．

MFCC，ΔMFCC，FBANK 全てにおいて局所特徴量を用いることで認識精度の改善が得られた．しかし，単一の特徴量として使える認識精度を得ることはできておらず，複数の特徴量を組み合わせて学習・認識を行うことで，認識精度の改善が期待できる．

Table 2　各特徴量での認識実験結果

特徴量	単語正解精度 [%]
MFCC	78.5
Proposed(MFCC)	81.5
ΔMFCC	49.2
Proposed(ΔMFCC)	52.6
FBANK	78.5
Proposed(FBANK)	82.1

4　おわりに

本稿では，時間-デルタケプストラム平面上の局所的な幾何学構造に基づき特徴を抽出する手法を用いて，構音障害者の音声認識実験を行った．提案手法によって，ΔMFCC において 3.4 %の改善が得られた．今後の課題として，複数の特徴量を組み合わせて認識精度の改善を行うことがあげられる．また，構音障害者特有の特徴量の検討や，様々な音声認識手法の適用を行い認識精度の改善に取り組んでいく．

参考文献

[1] 中村 他，“ 発話障害者音声を対象にした健常者音響モデルの適応と検証，” 音講論 (秋)，109-110，2005.

[2] H. Matsumasa *et al.*，“Integration of Metamodel and Acoustic Model for Speech Recognition,” Interspeech2008，2234-2237，2008.

[3] S. Terry Canale *et al.*，“キャンベル整形外科手術書 第 4 巻，” エルゼビア・ジャパン，2004.

[4] Y. Ariki *et al.*, “Phoneme Recognition Based on Fisher Weight Map to Higher-Order Local Auto-Correlation,” Interspeech2006, 377-380, 2006.

[5] T. Muroi *et al.*, “Speaker Independent Phoneme Recognition Based on Fisher Weight Map,” MUE2008, 253-257, 2008.

[6] 篠原 他, “フィッシャー重みマップを用いた顔画像からの表情認識,” 信学技報, PRMU2003-269, Vol.103, No.737, 79-84, 2004.